

## What Is Claimed Is:

1. A speech recognition system comprising:

5 microphone means for receiving input speech and converting the input speech into digital electronic signals;

10 front-end signal processing means, coupled to said microphone means, for processing the digital electronic signals to generate parametric representations of the digital electronic signals;

word database storage means for storing parametric representations of a plurality of predetermined word pronunciations;

15 word similarity comparator means, communicable with said front-end signal processing means and coupled to said word database storage means, for comparing the parametric representation of the digital electronic signals with said parametric representations of said plurality of predetermined word pronunciations, and generating a sequence of words selected from said plurality of predetermined word pronunciations in said word database storage means responsive to predetermined criteria for matching said input speech; and

20 display means, communicable with said word similarity comparator means, for displaying said sequence of words, comprising a scrolling means for scrolling through said sequence of words; a highlighting means, coupled to said scrolling means, for highlighting at least one of said sequence of words; and a selecting means, coupled to said scrolling means and said highlighting means, for selecting one of said sequence of words.

25 2. The speech recognition system of claim 1 wherein said front-end signal processing means generates digitally parameterized speech features for transmission to said word similarity comparator means.

30 3. The speech recognition system of claim 1 wherein said front-end signal processing means generates predetermined digitally parameterized acoustic parameters to said word similarity comparator means.

35 4. The speech recognition system of claim 1 wherein said front-end signal processing means comprises:

40 preemphasizer means for spectrally flattening the digital electronic signals generated by said microphone means;

45 frame-blocking means, coupled to said preemphasizer means, for blocking the digital electronic signals into frames of N samples with adjacent frames separated by M samples;

windowing means, coupled to said frame-blocking means, for windowing each frame;

autocorrelation means, coupled to said windowing means, for autocorrelating the frames;

cepstral coefficient generating means, coupled to said autocorrelation means, for converting each frame into cepstral coefficients; and

tapered windowing means, coupled to said cepstral coefficient generating means, for weighting the cepstral coefficients.

5            5. The speech recognition system of claim 1 wherein said word similarity comparator means comprises:

word calibration means, coupled to said word database storage means, for calibrating the parametric representations of the digital electronic signals with said parametric representations of said plurality of word pronunciation stored in said word database storage means;

dynamic time warper means for performing dynamic time warping on the parametric representations of the digital electronic signals and said parametric representations of said plurality of word pronunciations stored in said word database storage means;

distortion calculation means, coupled to said word calibration means and to said dynamic time warper means, for calculating a distortion between the parametric representations of the digital electronic signals and said parametric representations of said plurality of word pronunciations stored in said word database storage means;

scoring means, coupled to said distortion calculation means, for assigning a score to said distortion responsive to predetermined criteria; and

selection means, coupled to said scoring means, for selecting at least one of said parametric representations of said plurality of word pronunciations stored in said word database storage means having the lowest distortions.

6. The speech recognition system of claim 5 wherein said dynamic time warper means comprises minimization means for determining the minimum cepstral distances between the parametric representation of the digital electronic signals and said plurality of parametric representations of the word pronunciations stored in said word database storage means.

7. The speech recognition system of claim 6 wherein the words in said word database storage means corresponding to said selected at least one of said parametric representations of said plurality of word pronunciations stored in said word database storage means having the lowest distortions are displayed on said display in order of low to high distortions.

8. The speech recognition system of claim 1 wherein said microphone means, said front-end signal processing means, and said display are disposed in a mobile communication device.

9. The speech recognition system of claim 1 wherein said word database storage means, said word similarity comparator means are disposed in a server at a telecommunications site.

10. The speech recognition system of claim 8 wherein said mobile communication device communicates with the Internet.

11. The speech recognition system of claim 1 wherein said plurality of predetermined word pronunciation parametric representations include site category words on a communications network.

12. The speech recognition system of claim 1 wherein said plurality of predetermined word pronunciation parametric representations include company names on a stock exchange.

13. The speech recognition system of claim 1 wherein said plurality of predetermined word pronunciation parametric representations include transportation information related words.

14. The speech recognition system of claim 1 wherein said plurality of predetermined word pronunciation parametric representations include entertainment information related words.

15. The speech recognition system of claim 1 wherein said plurality of predetermined word pronunciation parametric representations includes restaurant information words.

16. The speech recognition system of claim 1 wherein said plurality of predetermined word pronunciation parametric representations include foreign exchange rate information related words.

17. The speech recognition system of claim 1 wherein said plurality of predetermined word pronunciation parametric representations include retail store name words.

18. The speech recognition system of claim 1 wherein said plurality of predetermined word pronunciation parametric representations include banking services related words.

19. The speech recognition system of claim 1 wherein said plurality of predetermined word pronunciation parametric representations include financial services related words.

20. The speech recognition system of claim 1 wherein said plurality of predetermined word pronunciation parametric representations include e-commerce and e-business related words.

21. A speech recognition system comprising:

microphone means for receiving a speech input and converting the speech input into digital electronic signals;

front-end signal processing means, coupled to said microphone means, for processing the digital electronic signals to generate parametric representations of the digital electronic signals;

word database storage means for storing a plurality of parametric representations of word pronunciations;

word similarity comparator means, communicable with said front-end signal processing means and to said word database storage means, for comparing the parametric representation of the digital electronic signals with said plurality of parametric representations of word pronunciations, and generating a first sequence of associations between the parametric representation of the digital electronic signals and said plurality of parametric representations of word pronunciations in said word database storage means responsive to predetermined criteria;

word string database storage means for storing a plurality of parametric representations of word string pronunciations;

word string similarity comparator means, coupled to said word similarity comparator and to said word string database storage means, for comparing a plurality of aggregated parametric representations of word pronunciations with said plurality of parametric representations of word string pronunciations in said vocabulary database storage means, and generating a second sequence of associations between at least one of said plurality of aggregated parametric representations of the word pronunciations with at least one of said plurality of parametric representations of word string pronunciations stored in said word string database storage means responsive to predetermined criteria; and

display means, coupled to said word string similarity comparator means, for displaying said second sequence of associations, comprising a scrolling means for scrolling through said second sequence of associations; a highlighting means, coupled to said scrolling means, for highlighting at least one of said second sequence of associations; and a selecting means, coupled to said scrolling means and said highlighting means, for selecting one of said second sequence of associations.

22. The speech recognition system of claim 21 wherein said word pronunciations and word string pronunciations are in the Korean language.

23. The speech recognition system of claim 21 wherein said word pronunciations and word string pronunciations are in the Japanese language.

24. The speech recognition system of claim 21 wherein said word pronunciations and word string pronunciations are in the Chinese language.

25. The speech recognition system of claim 21 wherein said microphone, said front-end signal processing means, and said display are disposed in a mobile communication device.

26. The speech recognition system of claim 25 wherein said mobile communications device communicates with the Internet.

27. The speech recognition system of claim 21 wherein said word database storage means, said word similarity comparator means, said word string database storage means, and said word string similarity comparator means are disposed in a server at a network communications site.

28. The speech recognition system of claim 21 wherein said plurality of word pronunciations and word string pronunciations include site category words on a communications network.

29. The speech recognition system of claim 21 wherein said plurality of word pronunciations and word string pronunciations include company names on a stock exchange.

30. The speech recognition system of claim 21 wherein said plurality of word pronunciations and word string pronunciations include transportation information related words.

31. The speech recognition system of claim 21 wherein said plurality of word pronunciations and word string pronunciations include entertainment information related words.

32. The speech recognition system of claim 21 wherein said plurality of word pronunciations and word string pronunciations include restaurant information words.

33. The speech recognition system of claim 21 wherein said plurality of word pronunciations and word string pronunciations include weather information words.

34. The speech recognition system of claim 21 wherein said plurality of word pronunciations and word string pronunciations include retail store name words.

35. The speech recognition system of claim 21 wherein said plurality of word pronunciations and word string pronunciations include banking services related words.

36. The speech recognition system of claim 21 wherein said plurality of word pronunciations and word string pronunciations include financial services related words.

37. The speech recognition system of claim 21 wherein said plurality of word pronunciations and word string pronunciations include e-commerce and e-business related words.

38. The speech recognition system of claim 21 wherein said plurality of word pronunciations and word string pronunciations include navigation aids words.

39. A speech similarity comparator comprising:

means for receiving digital electronic signals parametric representations;

word database storage means for storing a plurality of word pronunciation parametric representations;

word calibration means, coupled to said receiving means and to said word database storage means, for calibrating the digital electronic signals parametric representations with said plurality of word pronunciation parametric representations stored in said word database storage means;

dynamic time warper means for performing dynamic time warping on the digital electronic signals parametric representations and said plurality of word pronunciation parametric representations stored in said word database storage means;

distortion calculation means, coupled to said word calibration means and to said dynamic time warper means, for calculating a distortion between the digital electronic signals parametric representations and said plurality of word pronunciation parametric representations stored in said word database storage means;

scoring means, coupled to said distortion calculation means, for assigning a score to said distortion responsive to predetermined criteria;

word selection means, coupled to said scoring means, for selecting at least one of said plurality of word pronunciation parametric representations having the lowest distortion scores; and

display means, coupled to said selection means, for displaying said selected at least one of said plurality of words, said display means including a scrolling means for scrolling through said sequence of words; a highlighting means, coupled to said scrolling means, for highlighting at least one of said selected words; and a final selecting means, coupled to said scrolling means and said highlighting means, for a final selecting of one of said selected words.

40. The speech similarity comparator of claim 39 further comprising:

word string database storage means for storing a plurality of word string pronunciation parametric representations;

word string distortion calculation means, coupled to said word selection means, for calculating a distortion between an aggregation of said selected words parametric representations and said plurality of word string pronunciation parametric representations stored in said word string database storage means;

word string scoring means, coupled to said word string distortion calculation means, for assigning a score to said word string distortion responsive to predetermined criteria; and

word string selection means, coupled to said word string scoring means, for selecting at least one of said plurality of word string pronunciation parametric representations having the lowest word string distortion scores.

41. The speech similarity comparator of claim 40 wherein said display means, being communicable with said word string selection means, further displays said selected at least one of said plurality of word strings, scrolls through said selected word strings, highlights at least one of said selected word strings; and a finally selects one of said selected word strings.

42. In a network communications system having a plurality of mobile communication devices communicable with at least one server, and a plurality of sites communicable with the server, a speech recognition system comprising:

a signal processor, disposed in the mobile communication device, for digitally parameterizing a category word voice input signal;

a category word database, disposed in the server, for storing a plurality of predetermined parameterized category word pronunciations;

a category word comparator, communicable with said signal processor and coupled to said category word database, for comparing said digitally parameterized category word voice input signal with said plurality of parameterized category word pronunciations in said category word database;

a category word selector, coupled to said category word comparator, for selecting at least one category word of said plurality of parameterized category word pronunciations responsive to predetermined criteria utilized by said category word comparator; and

a display, coupled to said first selector, including a scrolling means for scrolling through said selected at least one category word; a highlighting means, coupled to said scrolling means, for highlighting at least one of said selected at least one category word; and a selecting means, coupled to said scrolling means and said highlighting means, for a final selecting of one of said selected at least one category word.

43. The speech recognition system of claim 42 further comprising a site map table, stored in said category word database, for mapping said selected category word to a category database stored in said category word database.

44. The speech recognition system of claim 42 wherein the server, responsive to said selection by said first selector selects at least one category word of said plurality of category word pronunciations, transmits said at least one category word back to said display for said final selection.

45. The speech recognition system of claim 42 further comprising

a speech words database, disposed at at least one of the plurality of sites, for storing a plurality of predetermined parameterized speech words pronunciations;

a speech words comparator, communicable with said signal processor and coupled to said speech words database, for comparing said digitally parameterized speech words voice input signal with said plurality of parameterized speech words pronunciations in said speech words database; and

a speech words selector, coupled to said speech words comparator, for selecting at least one speech word of said plurality of parameterized speech words pronunciations responsive to predetermined criteria utilized by said speech words comparator.

46. The speech recognition system of claim 45 wherein said speech words database includes a speech table for mapping said selected speech words to a predetermined content database.

47. The speech recognition system of claim 46 wherein said display further comprises content display means for displaying content from said predetermined content database.

48. The speech recognition system of claim 47 wherein at least one of said sites of said plurality of sites transmits content from said mapped predetermined content database to said display.

49. The speech recognition system of claim 45 wherein said speech words selector, responsive to said selection, transmits said at least one speech word back to said display for said final selection.

50. The speech recognition system of claim 42 wherein the network communication system is the Internet.

51. The speech recognition system of claim 42 where the plurality of sites are a plurality of websites.

52. In a network communications system having a plurality of mobile communication devices communicable with at least one server, and a plurality of sites communicable with the server, a speech recognition system comprising:

a signal processor, disposed in the mobile communication device, for digitally parameterizing a voice input signal;

a word database, disposed in the server, communicable with said signal processor, for storing a plurality of predetermined parameterized word pronunciations;



a word comparator, communicable with said signal processor and coupled to said word database, for comparing said digitally parameterized voice input signal with said plurality of parameterized word pronunciations in said word database;

5 a word selector, coupled to said word comparator, for selecting at least one of said plurality of parameterized word pronunciations in said word database responsive to predetermined criteria utilized by said word comparator;

10 a category word string database for storing a plurality of parameterized category word string pronunciations;

15 a category word string comparator, coupled to said signal processor and to said category word string database, for comparing an aggregation of said selected word pronunciations with said parameterized category word string pronunciations in said category word string database; and

20 a category word string selector, coupled to said category word string comparator, for selecting at least one word string from said plurality of parameterized category word string pronunciations in said category word string database responsive to predetermined criteria utilized by said category word string comparator.

25 53. The word recognition system of claim 52 further comprising a display, communicable with said category word string selector, for displaying said selected at least one word string from said plurality of category word string pronunciations, including a scroller for scrolling through said selection of word strings; a highlighter, coupled to said scrolling means, for highlighting at least one of said selection of word strings; and a final selector, coupled to said scroller and said highlighter, for selecting one of said highlighted word strings.

30 54. In a network communications system having a plurality of mobile communication devices communicable with at least one server, and a plurality of sites communicable with the server, a speech recognition system comprising:

35 a signal processor, disposed in the mobile communication device, for digitally parameterizing a voice input signal;

a speech word database, disposed in the server, communicable with said signal processor, for storing a plurality of parameterized speech pronunciations;

40 a speech word comparator, communicable with said signal processor and coupled to said speech word database, for comparing said digitally parameterized voice input signal with said plurality of parameterized speech word pronunciations in said speech word database;

45 a speech word selector, coupled to said speech word comparator, for selecting at least one of said plurality of parameterized word pronunciations in said speech word database responsive to predetermined criteria utilized by said speech word comparator;

a speech word string database for storing a plurality of parameterized speech word string pronunciations;

a speech word string comparator, coupled to said signal processor and to said database storage unit, for comparing said selected speech word pronunciations with said parameterized speech word string pronunciations in said parameterized speech word string database; and

a speech word string selector, coupled to said speech word string comparator, for selecting at least one of the word strings of said plurality of parameterized speech word string pronunciations in said speech word string database responsive to predetermined criteria utilized by said speech word string comparator.

55. The word recognition system of claim 54 further comprising a display, communicable with said speech word string selector, for displaying said selected at least one of the word strings of said plurality of speech word string pronunciations, including a scroller for scrolling through said selection of word strings; a highlighter, coupled to said scrolling means, for highlighting at least one of said selection of word strings; and a final selector, coupled to said scroller and said highlighter, for selecting one of said highlighted word strings.

56. In a network communications system having a plurality of mobile communication devices communicable with at least one server, and a plurality of sites communicable with the server, a speech recognition system comprising:

a signal processor, disposed in the mobile communication device, for digitally parameterizing a voice input signal;

a word database, disposed in the server, communicable with said signal processor, for storing a plurality of parameterized word pronunciations;

a word comparator, communicable with said signal processor and coupled to said word database, for comparing said digitally parameterized voice input signal with said plurality of parameterized word pronunciations in said word database;

a word selector, coupled to said word comparator, for selecting at least one of said plurality of parameterized word pronunciations in said word database responsive to predetermined criteria utilized by said word comparator;

a category word string database for storing a plurality of parameterized category word string pronunciations;

a category word string comparator, communicable with said signal processor and coupled to said category word string database, for comparing and aggregation of said selected words pronunciations with said category word string pronunciations in said category word string database;

1 a category word string selector, coupled to said category word string comparator, for  
selecting at least one of the word strings of said plurality of category word string  
pronunciations in said category word string database responsive to predetermined criteria  
utilized by said category word string comparator, thereby selecting at least one word  
5 string category;

1 a speech word database, disposed in the server, communicable with said signal  
processor, for storing a plurality of parameterized speech pronunciations;

10 a speech word comparator, communicable with said signal processor and coupled to  
said speech word database, for comparing said digitally parameterized voice input signal  
with said plurality of parameterized speech word pronunciations in said speech word  
database;

15 a speech word selector, coupled to said speech word comparator, for selecting at least  
one of said plurality of word pronunciations in said speech word database responsive to  
predetermined criteria utilized by said speech word comparator;

20 a speech word string database for storing a plurality of parameterized speech word  
string pronunciations;

25 a speech word string comparator, coupled to said signal processor and to said  
database storage unit, for comparing said selected speech word pronunciations with said  
speech word string pronunciations in said speech word string database;

30 a speech word string selector, coupled to said speech word string comparator, for  
selecting at least one word string from said plurality of parameterized speech word string  
pronunciations in said speech word string database responsive to predetermined criteria  
utilized by said speech word string comparator;

35 a plurality of content databases, respectively disposed at the plurality of sites,  
comprising information contents; and

40 a speech word string-to-content mapper for mapping said selected word string to at  
least one of said plurality of content databases.

57. The word recognition system of claim 56 further comprising a display,  
communicable with said category word string selector, for displaying said selected at  
least one category word string from said plurality of parameterized category word string  
45 pronunciations, including a scroller for scrolling through said selection of word strings; a  
highlighter, coupled to said scrolling means, for highlighting at least one of said selection  
of category word strings; and a final selector, coupled to said scroller and said highlighter,  
for selecting one of said highlighted category word strings.

58. The word recognition system of claim 56 further comprising a display,  
communicable with said speech word string selector, for displaying said selected at least  
one word string from said plurality of speech word string pronunciations, including a  
scroller for scrolling through said selection of word strings; a highlighter, coupled to said

scrolling means, for highlighting at least one of said selection of word strings; and a final selector, coupled to said scroller and said highlighter, for selecting one of said highlighted word strings.

59. The speech recognition system of claim 56 wherein the plurality of sites transmits said information contents to the mobile communication device responsive to said mapping.

60. A method for recognizing speech input, comprising the steps of:

- (a) parameterizing a predetermined plurality of word pronunciations;
- (b) storing said predetermined plurality of parameterized word pronunciations;
- (c) receiving the speech input;
- (d) converting the speech input into digital electronic signals;
- (e) parameterizing the digital electronic signals;
- (f) comparing said parameterized digital electronic signals with said stored predetermined plurality of parameterized word pronunciations;
- (g) selecting at least one of said stored predetermined plurality of parameterized word pronunciations responsive to predetermined parameter similarity criteria;
- (h) displaying said selected at least one of said stored plurality of parameterized word pronunciations;
- (i) scrolling through said selected at least one of said stored plurality of parameterized word pronunciations;
- (j) highlighting at least one of said selected at least one of said stored plurality of parameterized word pronunciations; and
- (k) further selecting one of said highlighted selected at least one of said stored plurality of parameterized word pronunciations.

61. The method of claim 60 wherein step (a) comprises utilizing cepstral coefficients to parameterize said plurality of parameterized word pronunciations.

62. The method of claim 60 wherein step (e) comprises utilizing linear predictive coding to parameterize the digital electronic signals.

63. The method of claim 60 wherein step (f) comprises utilizing cepstral distances to compare the parameterized digital electronic signals with said plurality of parameterized word pronunciations.

64. A method for recognizing speech input, comprising the steps of:

- (a) parameterizing a predetermined plurality of word pronunciations;
- (b) storing said predetermined plurality of parameterized word pronunciations;
- (c) parameterizing a predetermined plurality of word string pronunciations;
- (d) storing said predetermined plurality of parameterized of word string pronunciations;
- (e) receiving the speech input;
- (f) converting the speech input into digital electronic signals;
- (g) parameterizing the digital electronic signals;
- (h) comparing said parameterized digital electronic signals with said stored predetermined plurality of parameterized word pronunciations;

- (i) selecting at least one of said stored predetermined plurality of parameterized word pronunciations responsive to predetermined parameter similarity criteria;
- (j) aggregating said selected at least one of said stored plurality of parameterized word pronunciations to form a parameterized word string representing a word string;
- 5 (k) comparing said parameterized word string with said stored predetermined plurality of parameterized word string pronunciations;
- (l) selecting at least one of said stored predetermined plurality of parameterized word string pronunciations responsive to predetermined parameter similarity criteria;
- 10 (m) displaying said selected at least one of said word strings represented by said word string pronunciations in said stored plurality of parameterized word string pronunciations in a similarity sequence responsive to said predetermined parameter similarity criteria;
- (n) scrolling through said selected at least one of said stored predetermined plurality of parameterized word string pronunciations;
- 15 (o) highlighting at least one of said word strings in said selected at least one of said stored predetermined plurality of parameterized word string pronunciations; and
- (p) further selecting one of said highlighted selected at least one of said stored predetermined plurality of parameterized word string pronunciations.

20 65. The method of claim 64 wherein step (c) comprises utilizing cepstral coefficients to parameterize said predetermined plurality of parameterized word string pronunciations.

25 66. The method of claim 64 wherein step (k) comprises utilizing cepstral distances to compare said parameterized word string with said predetermined plurality of parameterized word string pronunciations.

30 67. A method of voice communication with a network communication system having a plurality of sites comprising the steps of:

- 35 (a) comparing a voice input comprising at least one category word with a database of predetermined category word pronunciations;
- (b) selecting at least one category word from said database of predetermined category word pronunciations responsive to predetermined word similarity criteria; and
- (c) establishing a communication link between the situs of said voice input and the site corresponding to said selected at least one category word responsive to said selected at least one category word.

40 68. The method of claim 67 further comprising the steps after step (b) of:  
displaying said selected at least one category word;  
scrolling through said selected at least one category word;  
highlighting at least one of said selected at least one said category word; and  
further selecting one of said highlighted selected at least one category word.

45 69. The method of claim 67 further comprising the steps of:

- (d) comparing a voice input comprising at least one speech word with a database of predetermined speech word pronunciations;

- (e) selecting at least one speech word from said database of predetermined speech word pronunciations; and
- (f) transmitting said selected at least one speech word to said site corresponding to said selected at least one category word.

70. The method of claim 69 further comprising the steps after step (e) of:

displaying said selected at least one speech word;  
scrolling through said selected at least one speech word;  
highlighting at least one of said selected at least one speech word; and  
further selecting one of said highlighted selected at least one speech word.

71. The method of claim 69 further comprising the step of:

(g) transmitting content from said site corresponding to said selected at least one category word responsive to said selected at least one speech word.

72. The method of claim 69 further comprising the steps after step (e) of:

aggregating said selected speech words to form word strings;  
comparing said speech word strings with a database of predetermined word string pronunciations;  
selecting at least one speech word string from said database of predetermined speech word string pronunciations; and  
transmitting said selected one speech word string pronunciation to said site corresponding to said selected at least one category word.

73. The method of claim 72 further comprising the steps, after the step of selecting at least one speech word string, of:

displaying said selected at least one speech word string;  
scrolling through said selected at least one speech word string;  
highlighting at least one of said selected at least one speech word string; and  
further selecting one of said highlighted selected at least one speech word string.

74. The method of claim 72 further comprising the step of:

transmitting content from said site corresponding to said selected at least one category word responsive to said selected at least one speech word string.